



INTRODUCCIÓN A LA CIENCIA DE DATOS Y EL BIG DATA

LOS DATOS

Hay miles de herramientas para analizar Big Data y para aplicar la Ciencia de Datos, sin embargo, no todas son iguales, pues hay que considerar que existen tres tipos de datos y que a continuación se describirán:

1. **Datos estructurados:** son aquellos donde los datos tienen una estructura bien particular, como son las fechas, los números, entre otros. Un ejemplo de ellos son las hojas de cálculo.
2. **Datos no estructurados:** usualmente son datos que tienen un formato específico y no se pueden almacenar en hoja de cálculos y mucho menos manipular la información, un ejemplo de esto son los documentos en PDF.
3. **Datos semiestructurados:** este tipo de datos no tienen un formato en particular, ya que cuentan con sus propios formatos denominados metadatos semi-estructurados, ejemplo de estos son los códigos de las páginas web (HTML).

Ahora bien, algo importante a destacar es que Big Data NO es una cosa; tampoco se trata del volumen; la variedad o la velocidad de los datos más de lo que las carreras de autos se relacionan con la gasolina.

Big Data se trata de ser más efectivo para aprovechar los datos y la analítica (análisis avanzado de datos) detrás de ellos para impulsar los modelos de negocios en una organización.

Los componentes técnicos (datos y analítica) para construir esa efectividad de Big Data para los negocios se describen a continuación:

Datos: Habilidad para recolectar y agregar datos detallados de una variedad de Fuentes incluyendo datos estructurados (archivos de excel; bases de datos; etc.), datos semi-estructurados (archivos de registro de transacciones, como las transacciones que se realizan con los bancos (log files); archivos de páginas web (XML files); archivos de intercambio de datos (JSON files); etc.) y los datos no estructurados (textos; videos; audios; imágenes, etc.)

Analítica: Habilidad para aprovechar el análisis avanzado de datos para descubrir información valiosa de clientes; productos; servicios; operaciones y mercados.

Entiendo que son los datos y que es la analítica, ahora cobra sentido el término Ciencia de Datos, más adelante se explicará con más detalle los análisis que forman parte de esta disciplina.

Big Data Business Model Maturity Index



Figura 1. Índice de Maduración del Modelo de Negocios de Big Data

En la figura 1 se muestra un modelo desarrollado por la organización Dell EMC en donde se define un índice de madurez para ayudar a una organización a medir el nivel de efectividad que obtuvieron al utilizar el análisis de datos y Big Data para el impulso de sus modelos de negocio. Este modelo es una especie de guía para que las empresas puedan tener una especie de "checklist" que les sirva para generar acciones.

En la fase "**Business monitoring**" las organizaciones están utilizando herramientas de Inteligencia de Negocios (Business Intelligence) y de almacenamiento de datos para supervisar el negocio; proporcionando una visión retrospectiva y por lotes de lo que ha logrado el negocio. La fase "**Insights**" trata

de combinar la riqueza de datos internos y externos con el análisis predictivo (e incluso el aprendizaje automático) para descubrir ideas sobre los procesos de negocio (internos) clave de la organización, atributos de productos y servicios, y / o comportamientos y sentimientos del cliente.



En la fase "**Optimization**" se aplica el análisis prescriptivo al cliente, al producto, a las operaciones y al mercado para ofrecer recomendaciones a los empleados, socios y clientes

y así mejorar la efectividad de los procesos de negocio clave (internos) de la organización.

La fase de "**Monetization**" aprovecha el enfoque de la fase "Optimization" para identificar y ejecutar nuevas oportunidades de negocio (externas) en el contexto de la estrategia comercial existente de la organización.

Finalmente, en la fase "**Metamorphosis**" se aprovechan los conocimientos acumulados de la organización, los datos y el análisis para crear nuevos componentes a la estrategia de negocios: nuevos modelos de negocios, nuevos modelos de consumo, nuevos objetivos corporativos, etc. En el siguiente curso se hablará con más detalle de este modelo.

Hasta este momento ha quedado claro que Big Data sirve de hilo conductor para que una organización pueda transformarse y así optimizar sus procesos y operaciones clave; reducir el riesgo de seguridad y cumplimiento; descubrir nuevas oportunidades de ingresos; y crear relaciones con el cliente más convincentes y diferenciadas.



Big Data trata de explotar las características únicas de los datos y de la analítica como activos digitales para crear nuevas fuentes de valor económico para una organización. La mayoría de los activos exhiben una relación transaccional de uno a uno.

Por ejemplo, el valor cuantificable de un dólar como activo es limitado: solo se puede usar para comprar un artículo o servicio a la vez. Lo mismo ocurre con los activos humanos, ya que una persona solo puede hacer un trabajo a la vez. Pero medir el valor de los datos como un activo no está limitado por esas limitaciones transaccionales. De hecho, los datos son un activo inusual ya que exhibe un Efecto Multiplicador Económico, por el cual nunca se agota o se desgasta y se puede usar simultáneamente en múltiples casos de uso con costo de margen cercano a cero. Esto hace que los datos sean un activo muy poderoso en el cual es muy atractivo invertir.

Entendiendo las características económicas de los datos y la analítica como activos digitales es el primer paso para monetizar los datos de una organización a través de análisis descriptivos; predictivos y prescriptivos. A continuación, se describirán cada uno de estos tipos de análisis; así como un breve resumen de que significa Business Intelligence (Inteligencia de Negocios) y Advanced Analytics (Analítica Avanzada) y cuál es su relación con Big Data.

¡De BBD a ABD... la evolución de los datos!



El cambio en la forma en que las empresas trabajan ha sido tan significativo en la última década que, de acuerdo con Harvard Business Review, las personas que han estado en el negocio durante al menos una década podrían definir sus carreras como BBD y ABD, o antes (Before) Big Data y después (After) Big Data.

Para ir comprendiendo como el análisis de los datos ha evolucionado hay que empezar hablando de la inteligencia de negocios (Business Intelligence); una disciplina que surgió en la década de 1950 cuando se desarrollaron herramientas que podían capturar información e identificar patrones y tendencias más rápido que la mente humana. Estos primeros esfuerzos en inteligencia de negocios representaron lo que los analistas de datos a menudo llaman **Analytics 1.0**.

Las características de la época incluyeron fuentes de datos pequeñas, estructuradas y en su mayoría internas, operaciones de procesamiento por lotes que podrían llevar meses e informes limitados y descriptivos (análisis descriptivo). Los analistas de ese tiempo dedicaban más tiempo a recopilar y preparar datos que a analizarlos realmente.



La era de la inteligencia de negocios duró alrededor de medio siglo, desde mediados de la década de 1950 hasta el 2009; en donde la era de la Big Data empezaba a gestarse.

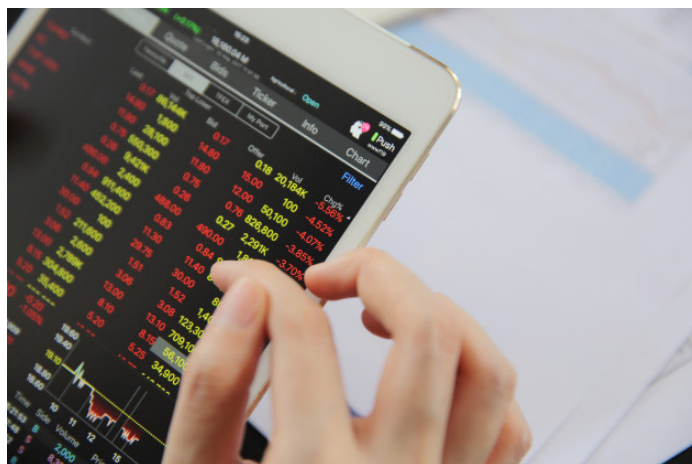
A mediados de la década de 2000, los gigantes de Internet y redes sociales como Google y Facebook comenzaron a descubrir, recopilar y analizar un nuevo tipo de datos. Si bien el término “Big Data” no entró en el léxico común hasta aproximadamente en el 2010, los analistas reconocieron que esta nueva información era cualitativamente diferente de los datos “pequeños” del pasado.

Los datos “pequeños” se generaban mediante las operaciones y transacciones internas de una empresa, pero estos nuevos datos se obtuvieron externamente, extraídos de Internet, fuentes de datos públicos y proyectos como el Proyecto del Genoma Humano. Esto significó el cambio a **Analytics 2.0**.

Con la llegada de Big Data, se desarrollaron nuevas tecnologías y procedimientos para ayudar a las compañías a convertir los datos en información y ganancias. Big Data requería nuevos marcos de procesamiento, como Hadoop (marco de referencia para el uso de grandes volúmenes de datos), y nuevas bases de datos, como NoSQL (base de datos para datos no estructurados), para almacenarla y manipularla. Los analistas de datos de **Analytics 2.0** requerían competencias tanto en el análisis como en la tecnología de la información, lo que ayudó a mejorar los avances tecnológicos venideros.

Muchos expertos creen que ha llegado una tercera era, Analytics 3.0, que se ve reflejada en los nuevos servicios orientados al cliente que utilizan análisis avanzado de datos (Advanced Analytics) para proporcionar una experiencia de usuario muy personalizada.

Los análisis revolucionarios en memoria o en base de datos ahora se combinan con métodos como análisis predictivo; análisis prescriptivo o aprendizaje automático (machine learning) para ofrecer información y resultados en tiempo real.



En esta nueva era, han surgido nuevas disciplinas analíticas para complementar el análisis descriptivo en la cartera de la analítica de datos. Los análisis predictivos y prescriptivos, que dan una idea de la probabilidad de que ocurra un evento en el futuro y recomiendan posibles cursos de acción, están emergiendo como herramientas clave para los ejecutivos y directivos de las organizaciones. Los análisis están disponibles para apoyar a una mejor toma de decisiones. Esta nueva era, seguramente no reflejará el final de la historia evolutiva. De hecho, presenta nuevos desafíos y oportunidades para las empresas y los analistas de datos por igual. Aquellos individuos que pueden capturar datos y organizarlos, así como analizarlos y utilizarlos para tomar mejores decisiones en el ambiente de los negocios, son y seguirán teniendo una gran demanda.

Queda claro que el análisis de datos partió de la era de la inteligencia de negocios hasta la era de Big Data y ahora es común escuchar términos como analítica de datos (Advanced Analytics); análisis descriptivo; análisis predictivo o análisis prescriptivo. A continuación, se explicarán cada uno de estos términos para entender la correlación que existe entre ellos y como forman parte de Big Data.

Advanced Analytics es el examen autónomo o semiautónomo de datos o contenido utilizando técnicas y herramientas sofisticadas, generalmente más allá de las que se utilizan en la **inteligencia de negocios** tradicional (BI); y tiene el objetivo de descubrir conocimientos más profundos, hacer predicciones o generar recomendaciones. Está dividido en 3 grandes análisis: descriptivo, predictivo y prescriptivo.

El **análisis descriptivo** ayuda a los analistas a comprender la realidad actual del negocio. Necesita comprender el contexto de los datos históricos para comprender la realidad actual de dónde se encuentra el negocio hoy en día. Este enfoque ayuda a una organización a responder preguntas tales como qué estilos de productos se están vendiendo mejor este trimestre en comparación con el último trimestre, y qué regiones están exhibiendo el crecimiento más alto o más bajo.

El **análisis predictivo** ayuda a anticipar cambios basados en la comprensión de los patrones y las anomalías dentro de esos datos. Con este modelo, el analista asimila una cantidad de fuentes de datos relacionadas para predecir los resultados. El análisis predictivo aprovecha los sofisticados algoritmos de **aprendizaje automático** (machine learning) para obtener información continua.

de los casos, el análisis prescriptivo predice no solo lo que sucederá, sino también por qué ocurrirá al proporcionar recomendaciones sobre acciones que aprovecharán las predicciones. Es como si la organización tuviera un médico de cabecera que estuviera monitoreando de forma constante las operaciones del negocio para “prescribir” una acción o decisión sobre una anomalía que detecte que va a suceder.



Finalmente, el relativamente nuevo campo del **análisis prescriptivo** permite a los analistas “prescribir” una serie de posibles acciones diferentes y guiarlos hacia una solución. En pocas palabras, estos análisis tienen que ver con proporcionar consejos.

El análisis prescriptivo intenta cuantificar el efecto de las decisiones futuras para asesorar sobre los posibles resultados antes de que las decisiones se tomen en realidad. En el mejor

Se prohíbe la reproducción total o parcial de esta obra por cualquier medio sin previo y expreso consentimiento por escrito del Instituto Tecnológico y de Estudios Superiores de Monterrey.

D.R. © Instituto Tecnológico y de Estudios Superiores de Monterrey, México. 2019 Ave. Eugenio Garza Sada 2501 Sur Col. Tecnológico C.P. 64849 Monterrey, Nuevo León | México



**Tecnológico
de Monterrey**